

Preprint No. M 13/09

A uniform concentration-of-measure
inequality for multivariate kernel
density estimators

Silvia Vogel, Anne Schettler

2013

Impressum:

Hrsg.: Leiter des Instituts für Mathematik
Weimarer Straße 25
98693 Ilmenau
Tel.: +49 3677 69-3621
Fax: +49 3677 69-3270
<http://www.tu-ilmenau.de/math/>

A uniform concentration-of-measure inequality for multivariate kernel density estimators

Silvia Vogel, Anne Schettler

Zusammenfassung

Confidence sets for modes or level sets of densities are usually derived from the asymptotic distribution of a suitable statistic. Mostly one does not have further information about how close the asymptotic distribution comes to the true distribution for a fixed sample size n . In order to derive conservative confidence sets for each sample size recently an approach was suggested that does not need full information about a distribution, but instead employs a quantified version of semi-convergence in probability of random sets. The application of this approach to modes or level sets of density functions requires uniform concentration-of-measure results for the density estimators. The aim of the present paper is to prove a result of that kind for the multivariate kernel density estimator. The inequality is also of own interest as it provides a conservative confidence band for the density function.

Keywords: kernel density estimator, uniform concentration-of-measure

1 Introduction

Density estimation with kernels has a long tradition in statistics. Kernel estimates are well investigated concerning consistency and asymptotic distribution, cf. Rosenblatt [11], Parzen [9], Müller [7], Silverman [12], Dony & Einmahl [3]. Knowledge of the asymptotic distribution enables the derivation of confidence sets, for instance for level sets, cf. Mason & Polonik [6]. Concentration-of measure results in the L_1 -setting have been derived by Devroye & Lugosi [2]. However, to the best of our knowledge, there is no paper which deals with uniform concentration of measure for these kind of estimates.

Uniform concentration-of-measure results are a main tool for the derivation of non-asymptotic confidence sets, so-called universal confidence sets, if random approximations to the true problem with suitable convergence properties are available (cf. Pflug [10] and Vogel [15], [16]). The approach is usually very easy to apply, once the needed convergence properties have been proved.

Suppose that we would like to investigate an unknown distribution represented by a density function f_0 . Assume that a density estimator f_n with the following property is available:

$$\sup_{n \in N} P\left(\sup_{x \in R^p} |f_n(x) - f_0(x)| \geq \beta_{n,\kappa}\right) \leq \mathcal{H}(\kappa). \quad (1)$$

Here $\mathcal{H}(\kappa)$ denotes a function with the property $\lim_{\kappa \rightarrow \infty} \mathcal{H}(\kappa) = 0$ and $(\beta_{n,\kappa})_{n \in N}$ denote sequences of positive real numbers with $\lim_{n \rightarrow \infty} \beta_{n,\kappa} = 0$ for each $\kappa > 0$.

This inequality can immediately be used to derive a universal confidence band for f_0 , but, more important, offers the possibility to obtain confidence sets for level sets, argmax sets etc. As an example we provide a simple approach for the derivation of a confidence area for a level set of the density function f_0 . To avoid additional technical considerations we assume that f_0 is u.s.c.

A conservative confidence set for the level set $M^c = \{x \in R^p : f_0(x) \geq c\}$ can be obtained as follows. Consider the random sets $M_n^{\beta_{n,\kappa}} = \{x \in R^p : f_n(x) \geq c - \beta_{n,\kappa}\}$, $n \in N$. If, for a fixed $n \in N$, there is an $x_n \in M^c$ which does not belong to $M_n^{\beta_{n,\kappa}}$, then $f_0(x_n) \geq c$, but $f_n(x_n) < c - \beta_{n,\kappa}$. Hence $f_n(x_n) - f_0(x_n) < -\beta_{n,\kappa}$ and the probability of the event $M^c \setminus M_n^{\beta_{n,\kappa}} \neq \emptyset$ can be bounded by $\mathcal{H}(\kappa)$. Consequently

$$\sup_{n \in N} P(M^c \setminus M_n^{\beta_{n,\kappa}} \neq \emptyset) \leq \mathcal{H}(\kappa).$$

In order to derive a confidence set for M^c with a prescribed level $1 - \eta$ one determines κ_0 such that $\mathcal{H}(\kappa_0) \leq \eta$. Then for each sample size n the set $M_n^{\beta_{n,\kappa_0}}$ covers the true level set M^c at least with probability $1 - \eta$. Note that no knowledge about the whole distribution or the asymptotic distribution is needed and a confidence set for each sample size n can be derived. (In fact, for this application we need only the weaker assertion $\sup_{n \in N} P(\inf_{x \in R^p} (f_n(x) - f_0(x)) \leq -\beta_{n,\kappa}) \leq \mathcal{H}(\kappa)$.)

G. Pflug [10] used this approach in a somewhat different framework to derive confidence sets for the single-valued solution set of an optimization problem. The method was further developed by S. Vogel and co-authors ([15], [16], [17], [13]). There are results for constraint sets, optimal values and solution sets of constrained optimization problems where the objective functions and the constraints can be approximated simultaneously. Also confidence sets of the form $M_n + B(0, \beta_{n,\kappa})$ are available, where M_n denotes the set under consideration for the approximate problem and $B(0, \beta_{n,\kappa})$ a ball with radius $\beta_{n,\kappa}$. Level sets can be regarded as constraint sets. Many estimators being the solution to random optimization problems which with increasing sample size converge to a deterministic problem, the approach can immediately be employed to derive confidence sets in parametric or nonparametric statistics.

In each case the approximation of the objective and/or constraint functions by random functions plays a crucial role. Sufficient conditions for these convergence assumptions have been proved for functions which are expectations with respect to an unknown probability measure where the measure is approximated by the empirical measure (Pflug ([10] and Vogel [16]). Univariate regression functions are dealt with by Sinotina & Vogel ([13],[14]).

The aim of the present paper is to prove a statement of the form (1) for a multivariate density kernel estimate. With this result it becomes possible to derive

universal conservative confidence sets, for instance for the level sets as shown above, but also for argmax sets and related sets. The main result is Theorem 2 at the end of section 2. It is proved in 3 steps in section 2. A discussion of the results concludes the paper.

2 A concentration-of-measure inequality

Let X_1, X_2, \dots be i.i.d. random vectors with values in R^p which have a density f_0 . We consider the kernel density estimator of f_0 based on X_1, \dots, X_n , $n \geq 1$,

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n^{1/p}}\right).$$

where K is a kernel and $h_n > 0$ is the bandwidth.

We will derive a convergence rate $\beta_{n,\kappa}$ and a tail behavior function \mathcal{H} such that

$$\sup_{n \in N} P(\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \geq \beta_{n,\kappa}) \leq \mathcal{H}(\kappa) \quad (2)$$

for $\mathcal{X} \subset R^p$.

We assume that the kernel K satisfies the following conditions:

$$\begin{aligned} (K1) \quad & \int_{R^p} |K(u)| du < \infty, \\ (K2) \quad & \int_{R^p} K(u) du = 1, \\ (K3) \quad & \int_{R^p} u_i K(u) du = 0 \quad \forall i = 1, \dots, p, \\ (K4) \quad & \sup_{x \in R^p} |K(x)| = C_1 < \infty. \end{aligned}$$

Note that, for instance, any symmetric density function satisfies the above conditions. (K4) implies that $\mathbb{E}[f_n(x)]$ exists.

A main tool for our investigations is the bounded difference inequality in the multivariate form. For the readers convenience we quote this inequality. For a proof see for instance [8] or [1].

Bounded Difference Inequality. Let X_1, \dots, X_n be i.i.d. random vectors with values in R^p and $g|R^p \rightarrow R^1$ a measurable function.

If $\forall i = 1, \dots, n$

$$\sup_{x_1, \dots, x_n, x'_i} |g(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq c_i$$

then

$$P(g(X_1, \dots, X_n) - \mathbb{E}g(X_1, \dots, X_n) \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

We proceed as follows. Obviously, $\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \leq T1_n + T2_n$ where

$$\begin{aligned} T1_n &:= \left| \sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| - \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \right] \right|, \\ T2_n &:= \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \right]. \end{aligned}$$

Firstly we investigate the random part $T1_n$.

Theorem 1. Let the conditions (K1), (K2), and (K4) be satisfied. Then for a kernel estimator f_n with bandwidth h_n and kernel K the following relation holds:

$$\mathbb{P} \left(\left| \sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| - \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \right] \right| \geq t \right) \leq 2e^{-\frac{t^2 n h_n^2}{2C_1^2}}.$$

Proof. Let $g(x_1, \dots, x_n) := \sup_{x \in \mathcal{X}} |\tilde{f}_n(x, x_1, \dots, x_n) - f_0(x)|$ where $\tilde{f}_n(x, x_1, \dots, x_n) := \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-x_i}{h_n^{1/p}}\right)$. Then we have

$$\begin{aligned} & \sup_{x_1, \dots, x_n \wedge x'_i \in A} |g(x_1, \dots, x_n) - g(x_1, \dots, x'_i, \dots, x_n)| \\ &= \sup_{x_1, \dots, x_n \wedge x'_i \in R^p} \left| \sup_{x \in \mathcal{X}} |\tilde{f}_n(x, x_1, \dots, x_n) - f_0(x)| - \sup_{x \in \mathcal{X}} |\tilde{f}_n(x, x_1, \dots, x'_i, \dots, x_n) - f_0(x)| \right| \\ &\leq \sup_{x_1, \dots, x_n \wedge x'_i \in R^p} \sup_{x \in \mathcal{X}} \left| \tilde{f}_n(x, x_1, \dots, x_i, \dots, x_n) - \tilde{f}_n(x, x_1, \dots, x'_i, \dots, x_n) \right| \\ &= \sup_{x_1, \dots, x_n \wedge x'_i \in R^p} \sup_{x \in \mathcal{X}} \left| \frac{1}{nh_n} \left(K\left(\frac{x-x_1}{h_n^{1/p}}\right) + \dots + K\left(\frac{x-x_1}{h_n^{1/p}}\right) + \dots + K\left(\frac{x-x_n}{h_n^{1/p}}\right) \right) \right. \\ &\quad \left. - \frac{1}{nh_n} \left(K\left(\frac{x-x_1}{h_n^{1/p}}\right) + \dots + K\left(\frac{x-x'_1}{h_n^{1/p}}\right) + \dots + K\left(\frac{x-x_n}{h_n^{1/p}}\right) \right) \right| \\ &= \sup_{x_1, \dots, x_n \wedge x'_i \in A} \sup_{x \in \mathcal{X}} \left| \frac{1}{nh_n} \left(K\left(\frac{x-x_i}{h_n^{1/p}}\right) - K\left(\frac{x-x'_i}{h_n^{1/p}}\right) \right) \right|. \end{aligned}$$

According to (K4) the kernel K is bounded, hence we obtain for each $x'_i \in R^p$

$$\left| K\left(\frac{x - x_i}{h_n^{1/p}}\right) - K\left(\frac{x - x'_i}{h_n^{1/p}}\right) \right| \leq 2C_1.$$

Consequently the assumption of the bounded difference inequality is satisfied with $c_i := \frac{1}{nh_n} 2C_1$, $i = 1, \dots, n$, and we obtain

$$\forall t > 0 : \quad \mathbb{P}(|g(x_1, \dots, x_n) - \mathbb{E}[g(x_1, \dots, x_n)]| \geq t) \leq 2e^{-\frac{t^2 nh_n^2}{2C_1^2}}. \quad \square$$

We see that in order to derive a useful tail behavior function from this theorem the assumption

$$(B1) \quad \lim_{n \rightarrow \infty} nh_n^2 = \infty$$

has to be imposed.

Now we turn to the deterministic part and make again use of the triangle inequality:

$$\mathbb{E}[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)|] \leq T2.1_n + T2.2_n$$

where

$$\begin{aligned} T2.1_n &:= \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]| \right], \\ T2.2_n &:= \sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)|. \end{aligned}$$

For the investigation of the term $T2.1_n$ we use the Fourier transform. Let k denote the Fourier transform of the kernel K :

$$k(u) := \int_{R^p} e^{iu^T y} K(y) dy \quad \forall u \in R^p.$$

Because of (K1) we have $\int_{R^p} |k(u)| du < \infty$. Hence we can employ the inversion formula and obtain $K(u) := (\frac{1}{2\pi})^p \int_{R^p} e^{-iu^T y} k(y) dy \quad \forall u \in R^p$. In an analogous way we obtain the Fourier transform of f_0 :

$$\phi(u) := \int_{R^p} e^{iu^T y} f_0(y) dy \quad \forall u \in R^p.$$

Consequently f_n can be rewritten in the following form:

$$\begin{aligned} f_n(x) &= \frac{1}{nh_n} \sum_{l=1}^n \left(\left(\frac{1}{2\pi} \right)^p \int_{R^p} e^{-i \left(\frac{x - X_l}{h_n^{1/p}} \right)^T y} k(y) dy \right) \\ &= \frac{1}{nh_n} \sum_{l=1}^n \left(\left(\frac{1}{2\pi} \right)^p h_n \int_{R^p} e^{-i(x - X_l)^T u} k(h_n^{1/p} u) du \right) \\ &= \left(\frac{1}{2\pi} \right)^p \int_{R^p} e^{-ix^T u} k(h_n^{1/p} u) \left(\frac{1}{n} \sum_{l=1}^n e^{iX_l^T u} \right) du \end{aligned}$$

With $\phi_n(u) := \frac{1}{n} \sum_{l=1}^n e^{iX_l^T u} \quad \forall u \in R^p$ we obtain

$$f_n(x) = \left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-ix^T u} k(h_n^{1/p} u) \phi_n(u) du \quad \forall x \in \mathcal{X}.$$

Lemma. Let (K1), (K2), and the following condition be satisfied:

$$(K5) \quad \int_{R^p} |k(u)| du = C_2 < \infty.$$

Then we have for a kernel estimator f_n with bandwidth h_n and kernel K

$$\mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]| \right] \leq \frac{C_2}{2\pi\sqrt{n}h_n}.$$

Proof. Because of Jensen's inequality we obtain

$$\mathbb{E}^2 \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]| \right] \leq \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right].$$

Now we use the Fourier transform and employ Fubini's theorem:

$$\begin{aligned} & \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right] \\ &= \mathbb{E} \left[\sup_{x \in \mathcal{X}} \left| \left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-iu^T x} k(h_n^{1/p} u) \phi_n(u) du - \mathbb{E} \left[\left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-iu^T x} k(h_n^{1/p} u) \phi_n(u) du \right] \right|^2 \right] \\ &= \mathbb{E} \left[\sup_{x \in \mathcal{X}} \left| \left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-iu^T x} k(h_n^{1/p} u) \phi_n(u, \omega) du - \left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-iu^T x} k(h_n^{1/p} u) \mathbb{E}[\phi_n(u)] du \right|^2 \right] \\ &= \mathbb{E} \left[\sup_{x \in \mathcal{X}} \left| \left(\frac{1}{2\pi}\right)^p \int_{R^p} e^{-iu^T x} k(h_n^{1/p} u) (\phi_n(u) - \mathbb{E}[\phi_n(u)]) du \right|^2 \right]. \end{aligned}$$

With $|e^{-iu^T x}| = 1$ we can conclude that

$$\mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right] \leq \mathbb{E} \left[\left| \left(\frac{1}{2\pi}\right)^p \int_{R^p} |k(h_n^{1/p} u)| |\phi_n(u) - \mathbb{E}[\phi_n(u)]| du \right|^2 \right].$$

For an integrable real-valued function $(u, \omega) \rightarrow \tilde{X}(u, \omega) =: X(u)$, $u \in R^p$, we have because of the Cauchy-Schwarz-inequality

$$\begin{aligned} \mathbb{E} \left[\left(\int_{R^p} X(u) du \right)^2 \right] &= \mathbb{E} \left[\left(\int_{R^p} X(u) du \right) \left(\int_{R^p} X(v) dv \right) \right] \\ &\leq \left(\int_{R^p} \mathbb{E}^{\frac{1}{2}}[X^2(u)] du \right)^2. \end{aligned}$$

Hence with $X(u) = \left| k(h_n^{1/p}u) \right| |\phi_n(u) - \mathbb{E}[\phi_n(u)]|$ we obtain

$$\mathbb{E}^{\frac{1}{2}} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right] \leq \left(\frac{1}{2\pi} \right)^p \int_{R^p} \mathbb{E}^{\frac{1}{2}} \left[\left| k(h_n^{1/p}u) \right|^2 |\phi_n(u) - \mathbb{E}[\phi_n(u)]|^2 \right] du.$$

Furthermore

$$\begin{aligned} & \mathbb{E}^{\frac{1}{2}} \left[\left| k(h_n^{1/p}u) \right|^2 |\phi_n(u) - \mathbb{E}[\phi_n(u)]|^2 \right] \\ &= \left| k(h_n^{1/p}u) \right| \mathbb{E}^{\frac{1}{2}} \left[|\phi_n(u) - \mathbb{E}[\phi_n(u)]|^2 \right] = \left| k(h_n^{1/p}u) \right| \sqrt{\text{var}(\phi_n(u))}, \end{aligned}$$

where $\text{var}(\phi_n(u)) = \frac{1}{n^2} \sum_{k=1}^n \text{var}(e^{iu^T X_k}) = \frac{1}{n} \text{var}(e^{iu^T X_1}) \leq \frac{1}{n}$.

Summarizing,

$$\begin{aligned} \mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]| \right] &\leq \mathbb{E}^{\frac{1}{2}} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right] \\ &\leq \left(\frac{1}{2\pi} \right)^p \int_{R^p} \left| k(h_n^{1/p}u) \right| \sqrt{\text{var}(\phi_n(u))} du \\ &\leq \left(\frac{1}{2\pi} \right)^p \frac{1}{\sqrt{n}} \int_{R^p} \left| k(h_n^{1/p}u) \right| du \\ &= \left(\frac{1}{2\pi} \right)^p \frac{1}{\sqrt{n}} \frac{1}{h_n} \int_{R^p} |k(y)| dy. \end{aligned}$$

With the assumption concerning k we obtain

$$\mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - \mathbb{E}[f_n(x)]| \right] \leq \frac{C_2}{(2\pi)^p \sqrt{n} h_n}. \quad \square$$

Now we consider T2.2_n. For the following approach we need additional conditions for f_0 and K .

Lemma. Let (K1), (K2), (K3), and the following conditions be satisfied:

(Vf) f_0 is in $C^2(R^p)$ and its partial derivatives of order 1 and 2 are bounded, especially $\forall i, j \in \{1, \dots, p\} : \sup_{x \in R^p} \left| \frac{\partial^2 f_0(x)}{\partial x_i \partial x_j} \right| \leq C_3 < \infty$,

(K6) $\int_{R^p} |u|^2 K(u) du = C_4 < \infty$.

Then we have for a kernel estimator f_n with bandwidth h_n und kernel K

$$\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| \leq \frac{C_3 C_4}{2} h_n^{2/p}.$$

Proof. The term $\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)|$ can be rewritten as follows:

$$\begin{aligned} \sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| &= \sup_{x \in \mathcal{X}} \left| \mathbb{E} \left[\frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x - X_i}{h_n^{1/p}} \right) \right] - f_0(x) \right| \\ &= \sup_{x \in \mathcal{X}} \left| \frac{1}{h_n} \mathbb{E} \left[K \left(\frac{x - X_i}{h_n^{1/p}} \right) \right] - f_0(x) \right| \\ &= \sup_{x \in \mathcal{X}} \left| \frac{1}{h_n} \int_{\mathbb{R}^p} K \left(\frac{x - y}{h_n^{1/p}} \right) f_0(y) dy - f_0(x) \right|. \end{aligned}$$

Now we change variables $\frac{x-y}{h_n^{1/p}} = u$ and exploit the properties of K :

$$\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| = \sup_{x \in \mathcal{X}} \left| \int_{\mathbb{R}^p} K(u) f_0(x - h_n^{1/p} u) du - f_0(x) \right|.$$

Because of $\int_{\mathbb{R}^p} K(u) du = 1$ we have $\int_{\mathbb{R}^p} K(u) f_0(x) du - f_0(x) = 0$ and hence

$$\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| = \sup_{x \in \mathcal{X}} \left| \int_{\mathbb{R}^p} K(u) (f_0(x - h_n^{1/p} u) - f_0(x)) du \right|.$$

The Taylor expansion of f_0 yields

$$f_0(x - h_n^{1/p} u) = f_0(x) - h_n^{1/p} u^T (f'_0(x)) + \frac{1}{2} h_n^{2/p} u^T H_0(\zeta_{x,u}) u$$

where $f'_0(x)$ denotes the gradient, H_0 denotes the Hessian of f_0 , and $\zeta_{x,u} \in (x - u h_n^{1/p}, x)$. Consequently

$$\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| = \sup_{x \in \mathcal{X}} \left| \int_{\mathbb{R}^p} K(u) (-h_n^{1/p} u^T (f'_0(x)) + \frac{1}{2} h_n^{2/p} u^T H_0(\zeta_{x,u}) u) du \right|.$$

Because of $\int_{\mathbb{R}^p} u_i K(u) du = 0$ we obtain $\sup_{x \in \mathcal{X}} |\mathbb{E}[f_n(x)] - f_0(x)| \leq \frac{1}{2} C_3 C_4 h_n^{2/p}$. \square

Hence we have

$$\mathbb{E} \left[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \right] \leq \frac{C_2}{(2\pi)^p \sqrt{n} h_n} + \frac{1}{2} C_3 C_4 h_n^{2/p}$$

and can summarize the results as follows:

Theorem 2. Assume that the conditions (K1) - (K6) and (Vf) are satisfied. Then

$$P(\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \geq \beta_{n,\kappa}) \leq \mathcal{H}(\kappa) \quad \forall n \in \mathbb{N}$$

where $\beta_{n,\kappa} = \frac{\kappa}{\sqrt{nh_n}} + \frac{C_2}{(2\pi)^p \sqrt{nh_n}} + \frac{1}{2}C_3C_4h_n^{2/p}$ and $\mathcal{H}(\kappa) = 2e^{\frac{-\kappa^2}{2C_1^2}}$.

Proof. Because of the triangle inequality we have

$$\begin{aligned} \sup_{n \in \mathbb{N}} P(\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| \geq \beta_{n,\kappa}) &\leq \sup_{n \in \mathbb{N}} P(T1_n + T2.1_n + T2.2_n \geq \beta_{n,\kappa}). \\ &\leq \sup_{n \in \mathbb{N}} P\left(T1_n + \frac{C_2}{(2\pi)^p \sqrt{nh_n}} + \frac{h_n^{2/p} C_3 C_4}{2} \geq \beta_{n,\kappa}\right) \\ &= \sup_{n \in \mathbb{N}} P\left(\left|\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)| - \mathbb{E}\left[\sup_{x \in \mathcal{X}} |f_n(x) - f_0(x)|\right]\right| \geq \frac{\kappa}{\sqrt{nh_n}}\right) \\ &\leq 2e^{\frac{-\kappa^2}{2C_1^2}}. \quad \square \end{aligned}$$

Acknowledgement. The authors would like to thank Wolfgang Polonik and Klaus Ziegler for valuable discussions about the topic.

Literatur

- [1] Bartlett, P. (2006). *Proof of the bounded differences inequality*. www.cs.berkeley.edu/~bartlett/.../bdddiff.pdf
- [2] Devroye, L. & Lugosi, G. (2001). *Combinatorial Methodes in Density Estimation*. Springer.
- [3] Dony, J. & Einmahl, U. (2009). *Uniform in bandwidth consistency of kernel estimators at a fixed point*. In: High Dimensional Probability V (C. Houdre et al., eds.), 308 - 325.
- [4] Dünnebier, A. (2008). *Konfidenzintervalle für Modalwerte*. Diploma thesis, TU Ilmenau.
- [5] Hoeffding, W. (1963). *Probability inequalities for sums of bounded random variables*. Journal of the American Statistical Association. 58, No. 301: 13 - 30.
- [6] Mason, D.M. & Polonik, W. (2009). *Asymptotic normality of plug-in level set estimates*. Ann. Appl. Probab., 19, No. 3: 1108–1142.
- [7] Müller, H.G. (1984). *Smooth optimum kernel estimators of densities, regression curves and modes*. Ann. Math. Statist., 12: 766 –774.
- [8] McDiarmid, C. (1989). *On the method of bounded differences*. In: Surveys in Combinatorics 1989 (J. Siemons, ed.), 148 –188, Cambridge Univ. Press.

- [9] Parzen, E. (1962). *On estimation of a probability density function and mode*. Ann. Math. Statist., 33: 1065 –1076.
- [10] G.C. Pflug, G.C. (2003). *Stochastic Optimization and Statistical Inference*. (In: Stochastic Programming, A. Ruszczyński, A. Shapiro, eds., Handbooks in Operations Research and Management Science, Vol. 10, Elsevier,) 427–482.
- [11] Rosenblatt, M. (1956). *Remarks on some non-parametric estimates of a density function*. Ann. Math. Statist., 27: 832 – 837.
- [12] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- [13] Sinotina, T. & Vogel, S. (2011). *Universal confidence sets for the mode of a regression function*. To appear in: IMA Journal of Management Mathematics.
- [14] Sinotina, T. & Vogel, S. (2011). *Concentration inequality for kernel regression estimators*. Preprint TU Ilmenau, 2011.
- [15] Vogel, S. (2008). *Universal confidence sets for solutions of optimization problems*. SIAM J. on Optimization, 19: 1467–1488.
- [16] Vogel, S. (2008). *Confidence sets and convergence of random functions*. In: Festschrift in Celebration of Prof. Dr. Wilfried Grecksch’s 60th Birthday, C. Tammer, F. Heyde (Hrsg.), Shaker-Verlag.
- [17] Vogel, S. (2011). *Universal confidence sets - Estimation and relaxation*. Preprint, TU Ilmenau.